

AD-A096 398

NORTH CAROLINA UNIV AT CHAPEL HILL

TWO EXAMPLES OF TRANSFORMATIONS WHEN THERE ARE POSSIBLE OUTLIER--ETC(U)

JAN 81 R J CARROLL

F/6 12/1

AFOSR-80-0080

AFOSR-TR-81-0123

NL

UNCLASSIFIED

1 of 1
AD-A096 398

END
DATE
FILED
4-81
DTIC

AD A 096398

AFOSR-TR- 81-0123

REF ID: A6512

(4)

Two Examples of Transformations When There are Possible Outliers

Raymond J. Carroll

University of North Carolina at Chapel Hill

and

National Heart, Lung, and Blood Institute

NTIC
LECTER
MAR 17 1981
D
A

AMS 1970 Subject Classifications: Primary 62E20, Secondary 62G35

Research supported by the U.S. Air Force Office of Scientific Research under contract AFOSR-80-0080.

DRG FILE COPY

Approved for public release;
distribution unlimited.

259500

81 3 16 050

| REPORT DOCUMENTATION PAGE | | READERSHIP SCHEDULE BEFORE CONTINUING FORM | |
|---|--|--|--|
| 1. REPORT NUMBER 18 | 2. GOVT ACCESSION NO. AD-A096 398 | 3. RECIPIENT'S DATA NUMBER | |
| 4. TITLE (and Subtitle) 6 TWO EXAMPLES OF TRANSFORMATIONS WHEN THERE ARE POSSIBLE OUTLIERS. | | 5. TYPE OF REPORT Technical Report | |
| 6. AUTHOR(s) 10 Raymond J. Carroll | | 7. CONTRACT OR GRANT NUMBER 15 AFOSR-80-0080 | |
| 8. PERFORMING ORGANIZATION NAME AND ADDRESS University of North Carolina at Chapel Hill and National Heart, Lung and Blood Institute | | 9. PROGRAM ELEMENT PROJECT NUMBER 16 11 11 | |
| 10. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, DC 20332 | | 11. REPORT DATE 11 January 1981 | |
| 12. NUMBER OF PAGES 12 18 | | 13. SECURITY CLASSIFICATION OF THIS REPORT UNCLASSIFIED | |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. DECLASSIFICATION OR ANNUAL REVIEW SCHEDULE | |
| 16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; Distribution Unlimited | | | |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 26, if different from Report) | | | |
| 18. SUPPLEMENTARY NOTES | | | |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Power transformations; robustness; asymptotic theory; likelihood ratio test; outliers | | | |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number) We apply the Box and Cox (1964) power transformation family and robust alternatives developed by Bickel and Doksum (1981) and Carroll (1980) to data sets given by John (178). The robust methods perform quite well compared to normal theory likelihood methods. | | | |

Summary

We apply the Box and Cox (1964) power transformation family and robust alternatives developed by Bickel and Doksum (1981) and Carroll (1980) to data sets given by John (1978). The robust methods perform quite well compared to normal theory likelihood methods.

Keywords and Phrases: POWER TRANSFORMATIONS; ROBUSTNESS; ASYMPTOTIC THEORY; LIKELIHOOD RATIO TESTS; OUTLIERS

AIR FORCE SYSTEMS INFORMATION SOURCEBOOK (AFSC)
NOTICE OF APPROVAL TO DDC
This technical document has been reviewed and is
approved for distribution under AFSC AF 100-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

Introduction

Our basic framework for transformation is the power family (Box and Cox (1964)); for some unknown λ ,

$$Y_i^{(\lambda)} = x_i \beta + \sigma \varepsilon_i, \quad i = 1, \dots, N,$$

where

$$\begin{aligned} Y^{(\lambda)} &= (Y^\lambda - 1)/\lambda \quad (\lambda \neq 0) \\ &= \log Y \quad (\lambda = 0). \end{aligned} \tag{1.1}$$

Here $\{x_i\}$ are $(1 \times p)$ design vectors, β is a $(p \times 1)$ regression parameter, σ is a scaling constant, and $\{\varepsilon_i\}$ are independently and identically distributed with mean zero and distribution F . Of course, we want F to be the standard normal distribution function Φ , but in general normality, linearity and heteroscedasticity may not be simultaneously attainable so we think of F as symmetric and almost normal.

Box and Cox (1964), Andrews (1971), Atkinson (1973), Bickel (unpublished) and Carroll (1980) have considered the problem of testing whether a given value λ_0 results in the model (1.1), i.e., they test

$$H_0: \lambda = \lambda_0. \tag{1.2}$$

Box and Cox proposed a likelihood ratio test, while Atkinson proposed a computationally simpler variant; both have good power properties when $F = \Phi$, but Carroll (1980) shows they are sensitive to outliers and have highly inflated test levels (Type I errors) when $F \neq \Phi$. The tests proposed by Andrews and Bickel hold the correct test levels when $F \neq \Phi$ but are not very powerful when $F = \Phi$.

Because the normal theory likelihood estimates are very sensitive to outliers, Bickel and Doksum (1981) and Carroll (1980) introduced robust methods. Let ρ be a (usually) convex function, $\psi = \rho'$ be odd and χ be an even function.

For a given λ define $\beta(\lambda)$ and $\sigma(\lambda)$ as the solutions to

$$\sum_1^N \psi(r_i(\lambda)) x_i = 0 \quad (1.3)$$

$$\begin{aligned} \sum_1^N x(r_i(\lambda)) &= 0 \\ r_i(\lambda) &= (y_i^{(\lambda)} - x_i \beta)/\sigma. \end{aligned} \quad (1.4)$$

One then minimizes the function

$$\ell(\lambda) = N \log \sigma(\lambda) + \sum_1^N \rho((y_i^{(\lambda)} - x_i \beta)/\sigma(\lambda)) - (\lambda - 1) \sum \log y_i. \quad (1.5)$$

When $\rho(x) = x^2/2 = \chi(x)$, we obtain the maximum likelihood estimates of the parameters (λ, β, σ) when $F = \phi$. In general, (1.5) is the likelihood when F has density proportional to $\exp(-\rho(x))$, and (1.3) - (1.4) lead to Huber's Proposal 2 (1973) for robust regression. Bickel and Doksum obtain the limiting distributions for the estimates (λ, β, σ) , showing that they have better robustness properties than the normal theory MLE. Other recent references are Carroll (1981b), (1981c), Carroll and Ruppert (1981), Doksum and Wong (1981) and Hernandez and Johnson (1981).

In this paper we apply the robust methods to the two data sets given by John (1978). John (1978) and Carroll (1981a) originally studied these data sets because both exhibit possible outliers; Carroll's (1981a) reanalysis is based on robust methods without transformation. In both data sets the responses are positive so that the simple model (1.1) is easy to apply. We focus primarily on estimating λ and testing whether it is a specified value, i.e., we test (1.2) for various λ_0 .

Carroll (1980) proposed testing (1.2) by treating the function $\ell(\lambda)$ in (1.5) as if it were a likelihood, rejecting $H_0: \lambda = \lambda_0$ if

$$L = 2(\ell(\lambda_0) - \ell(\lambda^*)) > c_\alpha \quad (1.6)$$

where c_α is the appropriate chi-square percentage point. For the choice

$$\begin{aligned} \psi(x) &= -\psi(-x) \\ &= x \quad 0 \leq x \leq k \\ &= k \quad x > k \end{aligned} \quad (1.7)$$

$$x(v) = \psi^2(v) - \int \psi^2(x) (2\pi)^{-1/2} \exp(-x^2/2) dx,$$

he found that such a test was somewhat of a compromise among those previously proposed; it has good power properties even when $F \neq \epsilon$, but its level varies and can be higher than desired, although it has an approximately correct level at the normal distribution and the problem of the level is not as severe as that for the normal theory likelihood ratio test.

One can study the general test statistic (1.6) by using the asymptotic theory of Bickel and Doksum (1981), who achieve major simplifications by letting $\sigma \rightarrow 0$ and $N \rightarrow \infty$ simultaneously. It turns out that one can prove the following

Result Define $x(y) = y_\cdot(y) - 1$, $r_i = r_i(\lambda^*)$

and

$$\hat{E}_\psi^1 = N^{-1} \sum_1^N r_i$$

$$\hat{E}_\psi^2 = N^{-1} \sum_1^N r_i^2.$$

Then as $N \rightarrow \infty$ and $\sigma \rightarrow 0$, under the hypothesis $H_0: \lambda = \lambda_0$, the statistic

$$\hat{L}^* = (\hat{E}_\psi^1) (\hat{E}_\psi^2)^{-1} L \quad (1.8)$$

has a chi-square distribution with one degree of freedom.

Details are given in the appendix. The statistic (1.8) is similar to one given by Schrader and Hettmansperger (1980). The choice

$$x(y) = y\psi(y) - 1 \quad (1.9)$$

is suggested by Bickel and Doksum, and we will use it throughout this paper. The result is of limited practical interest (see the example in the appendix), but at least it suggests a plausible choice for x .

2. Applications

In this section we apply the methods we have discussed to two data sets introduced by John (1978). Following Bickel and Doksum (1981), we set $x(x) = x\psi(x) - 1$ and we use the following three choices of ψ :

| | |
|------------|---|
| (MLE) | $\psi(x) = x$ |
| ("Huber") | $\psi(x)$ as in (1.7), $k = 2.0$ |
| ("Hampel") | $\psi(x) = -\psi(-x)$ |
| | $= x \quad 0 \geq x \geq a = 2.0$ |
| | $= a \quad a < x \leq b = 3.5$ |
| | $= a(c-x)/(c-b) \quad b < x \leq c = 5.0$ |
| | $= 0 \quad x > c$ |

We include the "Hampel" because the data sets have potential outliers and, as in Carroll (1980), the influence function of λ^* is not bounded if ψ is monotone. A word of caution about "Hampel" is in order. Because ψ is not monotone, convergence difficulties may arise. Hence in maximizing the function (1.5) with $\rho' = \psi$, we find the values of $\beta(\lambda)$ and $c(\lambda)$ by first solving for the "Huber" and then doing two iterations of the weighted least squares algorithm with the "Hampel" ψ . In all examples, the function $\ell(\lambda)$ attained a unique minimum on the interval $|\lambda| \leq 2.0$.

The first data set is particularly interesting. In the original scale of the data ($\lambda = 1$), both John (1978) and Carroll (1981a) conclude that the data point with response $Y = 14$ is an extreme outlier, but except for this point the normal linear model fits well. An acceptable analysis would thus estimate λ as somewhere near 1. As predicted by the influence function calculations in Carroll (1980), the "MLE" estimate for λ is much more sensitive to the outlier than the "Huber", which in turn is more sensitive than the "Hampel"; see Table 1 for details.

When we treat observation #11 with response $Y = 14$ as an outlier and replace it by John's suggested $Y = 62.33$, we obtain the results given in Table #2. All three methods give essentially the same answer now, and it seems reasonable to accept $H_0: \lambda = 1.0$ and to conclude that no transformation is really necessary. From a mechanical viewpoint, a combination of transformation and fitting using the "Hampel" ψ seems to give the best overall analysis. However, the best practice would be to use all three methods for the most revealing analysis.

In Table 3 we present estimates of λ and the test statistic for $H_0: \lambda = 1.0$ obtained by varying observation #11. It is interesting to note that "Hampel" is not insensitive to the changing observation, although we can always conclude that no transformation is really necessary.

For the second data set, all three methods indicate that logarithms would be an acceptable transformation (see Carroll (1981b) for a discussion of the value of moving the MLE of λ to an easily interpretable value).

These examples, the empirical work in Carroll (1981a) and substantial theoretical work as in Huber (1977) all point to the desirability of using robust methods in transforming and analyzing data, along, of course, with other standard tools.

Table 1

The first data set described by John (1978). The estimation methods are as described in (1.3) - (1.5), while the test statistic L^* is given by (1.8) - (1.9). These are the original data.

| | "MLE" | "Huber" | "Hampel" |
|-------------|-------|---------|----------|
| λ^* | 1.91 | 1.66 | 1.02 |
| $L^*(1.0)$ | 3.3 | 2.1 | 0.0 |
| $L^*(0.5)$ | 8.2 | 6.7 | 1.1 |
| $L^*(0.0)$ | 15.1 | 13.9 | 4.4 |
| $L^*(-.5)$ | 24.2 | 23.6 | 9.7 |
| $L^*(-1.0)$ | 35.1 | 35.8 | 16.7 |

Table 2

This is the first data set described by John, except that observation #11 ($Y = 14$) has been modified to $Y = 62.33$. See Table #1 for more details.

| | "MLE" | "Huber" | "Hampel" |
|-------------|-------|---------|----------|
| λ^* | 1.31 | 1.30 | 1.30 |
| $L^*(1.0)$ | 0.5 | 0.5 | 0.5 |
| $L^*(0.5)$ | 3.3 | 3.4 | 3.4 |
| $L^*(0.0)$ | 8.8 | 8.8 | 8.8 |
| $L^*(-0.5)$ | 16.5 | 16.2 | 16.2 |
| $L^*(-1.0)$ | 26.2 | 25.3 | 28.6 |

Table 3

Various values of λ^* and $L^*(1.0)$ for John's (1978) first data set when observation #11 is varied.

| Observation #11 | MLE | MLE $L^*(1.0)$ | "Huber" | "Huber" $L^*(1.0)$ | "Hampel" | "Hampel" $L^*(1.0)$ |
|-----------------|------|-------------------|---------|-----------------------|----------|------------------------|
| 14.00 | 1.91 | 3.34 | 1.66 | 2.10 | 1.02 | .00 |
| 18.83 | 1.77 | 1.98 | 1.55 | 1.28 | .70 | 1.73 |
| 23.67 | 1.58 | 1.02 | 1.43 | .72 | 1.20 | .20 |
| 33.31 | 1.30 | .30 | 1.28 | .28 | 1.28 | .28 |
| 43.00 | 1.30 | .35 | 1.32 | .38 | 1.31 | .38 |
| 52.67 | 1.41 | .71 | 1.43 | .76 | 1.43 | .76 |
| 62.33 | 1.31 | .48 | 1.30 | .50 | 1.30 | .50 |

Table 4

The second data set given by John (1978). See Table #1 for conventions.

| | "MLE" | "Huber" | "Hampel" |
|-------------|-------|---------|----------|
| λ^* | .11 | .15 | .15 |
| $L^*(1.0)$ | 21.45 | 19.69 | 15.06 |
| $L^*(0.5)$ | 4.32 | 3.37 | 2.25 |
| $L^*(0.0)$ | .36 | .53 | .53 |
| $L^*(-0.5)$ | 9.25 | 9.08 | 9.08 |
| $L^*(-1.0)$ | 25.95 | 25.51 | 25.49 |

ACKNOWLEDGEMENT

I wish to thank A.C. Atkinson for bringing my attention to the fact that John's data sets could be transformed and for suggesting the use of robust methods in this context.

REFERENCES

Andrews, D.F. (1971): A note on the selection of data transformations.
Biometrika 58, 249-254.

Atkinson, A.C. (1973): Testing transformations to normality. *J. Royal Statist. Soc. Series B*, 35, 473-479.

Bickel, P.J. and Doksum, K.A. (1981): An analysis of transformations revisited.
To appear in *J. Amer. Statist. Assoc.*

Box, G.E.P. and Cox, D.R. (1964): An analysis of transformations. *J. Royal Statist. Soc. Series B*, 26, 211-252.

Carroll, R.J. (1980): A robust method for testing transformations to achieve approximate normality. *J. Royal Statist. Soc. Series B*, 42, 71-78.

— (1981a): Robust methods for factorial experiments with outliers. To appear in *Applied Statistics*.

— (1981b): Power transformations when the choice of power is restricted to a finite set. Unpublished manuscript.

— (1981c): Tests for regression parameters in power transformation models. Unpublished manuscript.

— and Ruppert, D. (1980): Prediction and the power transformation family.
Tentatively accepted by *Biometrika*.

Doksum, K.A. and Wong, C.W. (1981): Statistical tests after transformations.
Unpublished manuscript.

Hernandez, F. and Johnson, R.A. (1981): The large sample behavior of transformations to normality. To appear in *J. Amer. Statist. Assoc.*

Huber, P.J. (1977): *Robust Statistical Procedures*. SIAM, Philadelphia.

Schrader, R.M. and Hettmansperger, T.P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika* 67, 93-102.

Appendix

Example: Consider regression through the origin with $\lambda = 0$, $\sigma = 1$:

$$\log Y_i = \beta X_i + \epsilon_i$$

$$\sum_1^N X_i = \sum_1^N X_i^3 = 0$$

$$\sum_1^N X_i^2 = N \quad \sum_1^N X_i^4 = \mu_4.$$

If one tests $H_0: \lambda = 0$ by the likelihood ratio test (which is just (1.8) with $\rho(x) = x^2/2$, $\chi(y) = y^2$), standard likelihood methods show that when H_0 is true,

$$L = L^* \rightarrow Zd,$$

where Z has a chi-square distribution with one degree of freedom and

$$d = (E\epsilon_1^6 - 4E\epsilon_1^4 + 4 + \beta^2(6E\epsilon_1^4 - 8) + \beta^4\mu_4)$$

$$\times (7E\epsilon_1^4/3 + 10\beta^2 + \beta^4\mu_4)^{-1}.$$

The constant $d = 1$ when $F = \phi$ and one can actually transform to a normal distribution, but in general $d \neq 1$ so that the test L^* does not always have the correct asymptotic level.

Bickel and Doksum (1981) study the asymptotic behavior of $(\lambda^*, \beta^*, \sigma^*)$ by letting $\sigma \rightarrow 0$ at a known rate as $N \rightarrow \infty$. Define

$$v_i = \frac{d}{d\lambda} Y_i^{(\lambda)} \quad u_i = \frac{d^2}{d\lambda^2} Y_i^{(\lambda)}$$

$$q_i = \lim_{\sigma \rightarrow 0} v_i.$$

$$B_{22} = (E\psi') N^{-1} \sum_1^N x_i' x_i$$

$$B_{12} = - (E\psi') N^{-1} \sum_1^N x_i q_i = B_{21}'$$

$$B_{13} = B_{31} = 0$$

$$B_{11} = - (E\psi') N^{-1} \sum q_i^2$$

$$B_{33} = E\epsilon_1 \chi(\epsilon_1)$$

Without stating the precise details, it suffices to state that they show that as $N \rightarrow \infty$, $\sigma \rightarrow 0$,

$$N^{\frac{1}{2}}((\lambda^*, \beta^*, \sigma^*) - (\lambda, \beta, \sigma)) / \sigma \quad (A1)$$

$$= N^{-\frac{1}{2}} \sum_1^N B^{-1} w_i + o_p(1),$$

where

$$B = (B_{ij})$$

and

$$w_i = (q_i \psi(\epsilon_i), x_i' \psi(\epsilon_i), \chi(\epsilon_i)).$$

Bickel and Doksum, Carroll and Ruppert (1980) and Carroll (1981b,c) discuss the interesting point outside the scope of this paper that (A1) means that β^* is asymptotically normally distributed with mean zero and covariance $(E\psi^2)S/(E\psi')^2N$,

$$S = N^{-1} \sum_1^N x_i' x_i + Q, \quad (A2)$$

and Q is positive semi-definite. This distribution is different from that when λ is known by the factor Q .

Define $(\hat{\beta}, \hat{\sigma})$ as the solutions to (1.3) - (1.4) using λ_0 , i.e., $\hat{\beta} = \beta(\lambda_0)$, $\hat{\sigma} = \sigma(\lambda_0)$. Detailed calculations based on (A1) show that when $H_0: \lambda_0 = \lambda$ is true,

$$N^{\frac{1}{2}}(\hat{\sigma}^* - \hat{\sigma})/\sigma^* \xrightarrow{P} 0 \quad (A3)$$

$$N^{\frac{1}{2}}(\lambda^* - \lambda_0)/\sigma \xrightarrow{D} N(0, \text{Var } = e), \quad (A4)$$

where

$$e = (\hat{E}\psi^2)(\hat{E}\psi')^{-1} \lim_{N \rightarrow \infty} [N^{-1} \sum_1^N q_i^2 - (N^{-1} \sum_1^N q_i x_i)^2]^{-1}.$$

We are now in a position to state

Theorem A. When $H_0: \lambda = \lambda_0$ is true, asymptotically as $N \rightarrow \infty$ and $\sigma \rightarrow 0$ the statistic

$$\hat{L}^{**} = (\hat{E}\psi')(\hat{E}\psi^2)^{-1}(\hat{L} + \hat{D}) \quad (A5)$$

is distributed as chi-square with one degree of freedom, where

$$r_i = r_i(\lambda^*)$$

$$D = 2N((\hat{\sigma} - \sigma^*)/\sigma^*)(N^{-1} \sum r_i \psi(r_i) - 1)$$

$$\hat{E}\psi' = N^{-1} \sum_1^N \hat{N}_{\psi'}(r_i)$$

$$\hat{E}\psi^2 = N^{-1} \sum_1^N \hat{N}_{\psi^2}(r_i).$$

When x is given by (1.7), the term D in (A5) is non-zero and can be of considerable importance. When x is given by (1.9), $D = 0$ and we obtain

$$L^* = L^{**}.$$

Of course when $\rho(x) = x^2/2$, $\psi(x) = x$ and $\chi(y) = y^2 - 1$ we have that $L = L^* = L^{**}$, the normal theory likelihood ratio test. The example shows that the result stated in the body of the paper depends for its validity on the assumption that $\sigma \rightarrow 0$.

Proof of Therem A. The proof is based upon the following Lemma, which is extremely messy to obtain but only used Taylor expansions.

Lemma A. As $N \rightarrow \infty$ we have

$$L = (D_1 + D_2 + D_3 + D_4 + D_5 + D_6) \xrightarrow{P} 0,$$

where

$$D_1 = 2N((\hat{\sigma} - \sigma^*)/\sigma^*)(N^{-1}\sum r_i(\lambda^*)\psi(r_i(\lambda^*)) - 1)$$

$$D_2 = N((\lambda^* - \lambda)/\sigma)^2(N^{-1}\sum(\sigma u_i\psi(\varepsilon_i) + v_i^2\psi'(\varepsilon_i)) - (N^{-1}\sum v_i x_i)^2)$$

$$D_3 = 2((\sigma^* - \hat{\sigma})/\sigma)\sum(\varepsilon_i\psi(\varepsilon_i) - r_i(\lambda^*)\psi(r_i(\lambda^*)))$$

$$D_4 = -N((\sigma^* - \hat{\sigma})/\sigma)^2$$

$$D_5 = 2N\mathbb{E}\varepsilon_1\psi(\varepsilon_1)(\sigma^* - \hat{\sigma})(\sigma^* - \sigma)/\sigma^2$$

$$D_6 = N(((\hat{\sigma} - \sigma)/\sigma)^2 - ((\sigma^* - \sigma)/\sigma)^2)(\mathbb{E}\varepsilon_1^2\psi'(\varepsilon_1) + 2\mathbb{E}\varepsilon_1\psi(\varepsilon_1)).$$

Theorem A follows from Lemma A because of (A3) and (A4).

